

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/EP05/050409

International filing date: 01 February 2005 (01.02.2005)

Document type: Certified copy of priority document

Document details: Country/Office: DE
Number: 10 2004 008 192.1
Filing date: 18 February 2004 (18.02.2004)

Date of receipt at the International Bureau: 23 February 2005 (23.02.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

BUNDESREPUBLIK DEUTSCHLAND



Prioritätsbescheinigung über die Einreichung einer Patentanmeldung

Aktenzeichen: 10 2004 008 192.1

Anmeldetag: 18. Februar 2004

Anmelder/Inhaber: Siemens Aktiengesellschaft, 80333 München/DE

Bezeichnung: Patientenidentifikation zur Auswahl für Studien

IPC: G 06 F 17/30

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Patentanmeldung.

München, den 31. Januar 2005
Deutsches Patent- und Markenamt
Der Präsident
 Im Auftrag

Stanschus

Beschreibung

Patientenidentifikation zur Auswahl für Studien

- 5 1. Für die Patienten-Identifikation werden mehr und detailliertere Daten benötigt als momentan zugänglich sind - wobei implizit bereits viele Informationen abgespeichert sind (in Form von unstrukturierten Texten, in Form von Bildern, etc.).
- 10 Es geht also darum, bereits vorhandene, aber noch nicht zugängliche (z.B. aufgrund technologischer Hürden) Daten zu dem Zweck der Patienten- / Site- Identifizierung zugänglich und nutzbar zu machen.
- 15 Die Rekrutierung von geeigneten Probanden oder Patienten stellt den vielleicht aufwändigsten und problematischsten Schritt im Ablauf einer klinischen Studie dar. Bevor Patienten rekrutiert werden können, sind sie als ,studientauglich' zu identifizieren. Dies geschieht anhand
- 20 von Ein- und Ausschlusskriterien. Heute wird diese Bewertung vorwiegend manuell anhand von Papierunterlagen vorgenommen. Mit Zunahme elektronisch verfügbarer medizinischer Daten werden diese Aufgaben durch Suchalgorithmen in Datenbanken ersetzt. Die einfachste Suche kann in hochstrukturierten Daten erfolgen. Leider ist der Hauptteil medizinischer Einträge unstrukturiert, z.B. in Textform. Damit wird eine einfache Suche anhand von Stichwörtern aufwändig und bezüglich des Suchergebnisses unbefriedigend, da die Medizin keine standardisierte Terminologie verwendet. So kann z.B.
- 30 der ,zuckerkrank' identifiziert werden, der Diabetiker aber nicht, wenn einfache Suchalgorithmen verwendet werden. Noch komplizierter wird die Identifizierung von Patienten aus elektronischen Bildarchiven, z.B. die Erkennung von Tumoren in MR oder CT Bildern.

35

2. Suche in Papierakten manuell

Einfache Suche in strukturierten Daten mittels SQL o.ä.

Einfache Textsuche a la Google

5

3. Vorhandene Datenquellen (unstrukturierte Textdaten, Bilder, *omics daten) werden:

- nach Relevanz für Patienten- /Site-Identifizierung bewertet und identifiziert

10 - digitalisiert, soweit notwendig

- zugreifbar gemacht (Implementierung von Datenbank-Technologie und Erstellen einer Integrations-/Transformations-Zwischenschicht¹)

15

Unstrukturierte medizinische Daten

Nutzung von Domänenwissen in Form von Nomenklaturen und Klassifikationen, um auf Zugehörigkeit zu Krankheitsgruppen, spezifischen Krankheiten zu schließen

20 Nutzung von Domänenwissen in Form von Wissensbasen oder Expertensystemen, um Zugehörigkeit zu bestimmten Krankheiten implizit zu schließen.

Beispiele:

o Expertenwissen:

Eine Blutglukosekonzentration von 300mg/100ml legt den Schluß nahe, dass es sich um einen Diabetiker handelt

o Bayes-Klassifizierung von unstrukturiertem Text:

Vor allem wenn auf eine Vielzahl ähnlicher Dokumente zugreifbar ist, könnte eine Klassifikation der Dokumente nach dem Krankheitsbild, Behandlung etc. erfolgen.

30

Nach der Häufigkeit von im Text vorkommenden Schlüsselwörtern wird das Dokument einer (aus mehreren) Klassen zugeordnet. Der Schlüsselwortindex wird während einer Trainingsphase anhand von Beispieldokumenten

¹ Diese Schicht integriert die verschiedenen Datenquellen, sie stellt gegebenenfalls auch eine Verbindung zu Referenz-DBs (intern/extern) her und kann auch das Domänenwissen nutzbar machen.

- erstellt und kann laufend erweitert/verbessert werden.
Kann bei einem Teil der Datensätze durch zusätzliche,
aus strukturierten Daten gewonnenen Merkmalen vorab
eindeutig bestimmt werden, welcher Klasse der "Text"
5 zuzuordnen ist, könnte mit diesen Datensätzen die
Trainingsphase automatisiert werden.

Bilddaten

- 10 Erkennung des Bildinhaltes durch Vergleich mit
Referenzbilddatenbank oder Domänenwissen in anderer Form ->
Entscheidungsunterstützung über Zugehörigkeit zu
Patientengruppe
Patientenidentifizierung durch Kombination Bildinhalt -
15 medizinische Datenbank

Omics - Daten

- Patientenidentifizierung anhand von Kombinationen omics Daten
- klinische Daten - Bilddaten
20 Entscheidung anhand Vergleich Individualfall mit entsprechend
qualifizierter Datenbank

Anwendungsgebiete:

- o Patientenidentifizierung für:
 - i. Klinische Studien
 - ii. Zugehörigkeit zu Risikogruppen (entwickelt
Krebs etc.)
 - iii. Könnte nach Guideline A behandelt werden

30

4. Der erfinderische Schritt liegt darin, zum Zweck der
Patienten- /Site-Identifizierung Datenquellen zu nutzen, die
bisher noch nicht genutzt wurden (images, unstrukturierte
Texte, *omics).

- 35 Tools und Methoden werden zur Verfügung gestellt, die durch
die Einbindung von Domänenwissen in Form von
Referenzdatenbanken, qualifizierten klinischen Datenbanken,

200402288

omics Daten separat und in der Integration all dieser eine Patientenidentifizierung möglich machen, die bisher in dieser Form nicht möglich ist .

5

